

0007975-0016

**CERTIFICATE OF MAILING 37 CFR §1.10**

"Express Mail" Mailing Label Number EK 919497269 US

Date of Deposit. November 8, 2001

I hereby certify that this paper, accompanying documents and fee are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR §1.10 on the date indicated above and is addressed to Commissioner for Patents, Box Patent Application, Washington, D C 20231



Jose Ramos

**UNITED STATES PATENT APPLICATION**

**FOR**

**IMAGE AND VIDEO INDEXING SCHEME FOR  
CONTENT ANALYSIS**

**INVENTORS:**

PATRICIA A. KEATON  
RODNEY M. GOODMAN

**PREPARED BY:**

COUDERT BROTHERS LLP  
333 S. Hope Street, 23rd Floor  
Los Angeles, California 90071  
(213) 229-2900

This application claims priority of U. S. Provisional Application No. 60/247,469, filed on November 8, 2000. This provisional patent application is hereby incorporated herein by reference.

5

## BACKGROUND OF THE INVENTION

### 1. FIELD OF THE INVENTION

The present invention relates to image and video storage and retrieval systems.

10

Portions of the disclosure of this patent document contain material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the Patent and Trademark Office file or records, but otherwise reserves all copyright rights whatsoever.

15

### 2. BACKGROUND ART

Computer systems are used to store large amounts of information and data. To be useful, it is important that the data be organized and searchable so that data and information can be easily found. For text data it is relatively easy to search for data by searching for key words that might be found in the text of stored documents. Thus, the stored data itself can be used as part of the searching effort. It is not as easy to search for images on computer systems because of the way that they are stored. Images are stored, in one example, as a series of pixels that indicate a particular color. There is nothing about the pixel that lets a searcher know if it is a

part of a picture of a car or a bird. Thus, the image data itself has not been easily usable as part of the searching effort.

One method for making it easier to search for images is the use of captions or text descriptions associated with the image that themselves are searchable. For example, a picture of a car on a bridge could have a caption describing the scene with the car, bridge, background, etc. all described in text. When a person searches for an image, the person enters words that are then used to search through image captions. This scheme requires that each image be looked at and described by a human operator, a time consuming effort and one that adds to the amount of data needed to be stored with each image, so it is space consuming as well. This type of system is called a content-based retrieval system.

Another type of image and video storage and retrieval system uses a compressed domain approach. The compressed domain approach derives the image or video features from the transform coefficients, thus requiring decompression.

The problems associated with image indexing and retrieval systems can be better understood by a review of content-based retrieval systems and compressed domain systems.

20 Content Based Retrieval – Keyword Approach

One type of content based retrieval system uses keywords. Typically, keywords describing each image are recorded in text and associated with the image. (This additional data, which in part describes the image, is often referred to as “meta-data”). When a user wishes to

retrieve the image, a keyword is typed and all of the images having that associated keyword are retrieved. This requires great human effort in creating the meta-data that enables visual queries. The text descriptions also do not completely or consistently characterize the content of the images and videos. Second, the relatively large data sizes of images and videos compared to the communication channel bandwidth prohibits the user from browsing or perusing all but a small portion of the archive at a time. Therefore, the ability to find desired images and videos depends primarily on the capabilities of the query tools provided by the system.

#### Content Based Retrieval – Query Approach

10

Using a content-based query, the user provides a description of some of the prominent visual features of an image or video. Then, a mechanism is enabled by which the computer searches the archive and returns the images and videos that best match the description. Typically, research on content-based queries have focused on the visual features of color, texture and shape. For example, the IBM Query By Image Content (QBIC) project proposes and utilizes feature sets that capture the color, texture and shape of image objects that have been segmented manually. Texture and color features are also utilized that describe the global features of images.

20 The keyword based and query based approaches to content based retrieval store the keywords or visual features in addition to the compressed imagery. This produces a data expansion, which is disadvantageous.

#### Compressed Domain Retrieval

The advent of compression standards has led to the proliferation of indexing techniques in the compressed domain. Many images and videos in a networked multimedia database are of a compressed nature. Compressed domain techniques seek to identify and retrieve the images by processing data in the compressed representation of the images. The main advantage of  
5 compressed domain processing is the reduction of computational complexity which results from the smaller size of the compressed data file.

Compressed domain techniques, however, derive the features of the images or videos from their transform coefficients. This requires the decompression of the bit-stream up to an  
10 inverse transformation step, which is disadvantageous. There is currently no approach that minimizes the data expansion associated with content based retrieval and also minimizing the decompression associated with compressed domain approaches.

SUMMARY OF THE INVENTION

The present invention provides an image and video indexing scheme for content analysis. According to the invention, a database of images or videos is compressed. By examining patterns in the compression scheme of each image or video, the present invention identifies the content of the data. In one embodiment, an unsupervised learning method is employed where each image or video is sub-divided into smaller blocks (8 pixels x 8 pixels, for instance) and each of the smaller blocks is examined for its compression pattern. Then, the pattern associated with each of the smaller blocks is recorded for each of the images in the database and content is retrieved from the database by associating certain patterns or groups of patterns with certain content.

In one embodiment, the compression patterns for each of the subdivided blocks of the images or videos comprise a collection of transform codes. The transform codes are the codes used to compress the block into the fewest bits possible. The transformed codes are obtained via the unsupervised learning method that takes place across a database where the images and videos reside. The collection of transform codes is used to build a dictionary that provides a probabilistic description of the data that is suitable for content-based retrieval. The bit-stream is organized to support its progressive decoding.

20

In one embodiment, the unsupervised learning method finds the transform codes for the dictionary based upon latent variable modeling, which is adopted to learn a collection, or mixture, of local linear subspaces over a set of image patches or blocks. In one embodiment, Gaussian latent variable modeling defines a proper probability density model over the input

space, and concurrently performs a data partitioning and a reduction step within a maximum likelihood framework.

- In one embodiment, a user is able to query the database of images and videos by either
- 5 submitting an input image for retrieval or specifying a type of content to be retrieved. If the type of content to be retrieved is specified, then images with sets of compression patterns associated with the content are retrieved. If an input image is submitted, a set of compression patterns (e.g., transform codes) is obtained for the input image and matched with the sets of transform codes for the images in the database. To determine if the transform codes for the
  - 10 input and database images match, an optimal Bayes decision rule is used by one embodiment of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects and advantages of the present invention will become better understood with regard to the following description, appended claims and accompanying drawings where:

Figure 1 is a flow diagram illustrating one embodiment of the present invention.

Figure 2 is a flow chart illustrating the operation of one embodiment of the invention.

Figure 3 is a flow diagram illustrating progressive decoding.

Figure 4 is a flow diagram illustrating data partitioning.

Figure 5 is a flow diagram illustrating image retrieval.

Figure 6 is a block diagram of an embodiment of the present invention that uses an image-based user query.

Figure 7 is a flowchart showing a process according to an embodiment of the present invention.

Figure 8 is a diagram illustrating Bayesian evidential reasoning.

Figure 9 is a block diagram of a system for object detection.

Figure 10 is a block diagram of a general purpose computing environment.

## DETAILED DESCRIPTION OF THE INVENTION

The invention relates to an image and video indexing scheme for content analysis. In the following description, numerous specific details are set forth to provide a more thorough 5 description of embodiments of the invention. It is apparent, however, to one skilled in the art, that the invention may be practiced without these specific details. In other instances, well known features have not been described in detail so as not to obscure the invention.

### Image and Video Indexing Scheme

10

According to the invention, a database of images or videos is compressed. By examining patterns in the compression scheme of each image or video, the present invention identifies the content of the data. Figure 1 is a flowchart showing this embodiment of the present invention. At block 100, a database of images and videos are obtained. At block 110, the images and 15 videos are compressed. At block 130, patterns in the compression of each image or video are examined using an unsupervised learning method. At block 140, content is retrieved from the database using the compression patterns.

### Unsupervised Learning Method

20

In one embodiment, an unsupervised learning method is employed for is used to examine patterns in the source image. In the method of this embodiment, each image or video is sub-divided into smaller blocks (8 pixels x 8 pixels, for instance) and each of the smaller blocks is examined for its compression pattern. Then, the patterns associated with each of the

smaller blocks is recorded for each of the images in the database and content is retrieved from the database from the database by associating certain patterns or groups of patterns with certain content.

5       Figure 2 is a flowchart showing this embodiment of the present invention. At block 200, an image or video is obtained from the database. At block 210, the image is sub-divided into smaller blocks (8 pixels x 8 pixels, for instance). At block 220, each of the smaller blocks is examined for its compression pattern. At block 230, the patterns associated with each of the smaller blocks is recorded. At block 240, it is determined if all of the images or videos in the  
10      database have been learned. If not, the process repeats at block 200. Otherwise, content is retrieved from the database at block 250 by associating certain patterns or groups of patterns with certain content.

15      In one embodiment, the compression patterns for each of the subdivided blocks of the images or videos comprise a collection of transform codes. The transform codes are the codes used to compress the block into the fewest bits possible. The transform codes are obtained via the unsupervised learning method that takes place across a database where the images and videos reside. The collection of transform codes are used to build a dictionary that provides a probabilistic description of the data that is suitable for content-based retrieval. The bit-stream is  
20      organized to support its progressive decoding.

Figure 3 is a flowchart showing this embodiment of the present invention. At block 300, an image or video is obtained from the database. At block 310, the image is sub-divided into smaller blocks. At block 320, each of the smaller blocks is examined for the transform code

used to optimally compress the block. At block 330, the learned transform codes are added to a dictionary. At block 340, it is determined if all of the images or videos in the database have been learned. If not, the process repeats at block 300. Otherwise, content is retrieved from the database at block 350 by associating certain transform codes or groups of transform codes with  
5 certain content.

#### Latent Variable Modeling

In one embodiment, the unsupervised learning method finds the transform codes for the  
10 dictionary based upon latent variable modeling, which is adopted to learn a collection, or mixture, of local linear subspaces over a set of image patches or blocks. In one embodiment, Gaussian latent variable modeling defines a proper probability density model over the input space, and concurrently performs a data partitioning and a reduction step within a maximum likelihood framework.  
15

Figure 4 is a flowchart showing this embodiment of the present invention. At block 400, an image or video is obtained from the database. At block 410, the image data is partitioned into smaller blocks. At block 420, each of the smaller blocks has Gaussian latent variable modeling applied to them to obtain the transform code used to optimally compress the block. At block 20 430, the learned transform codes are added to a dictionary. At block 440, it is determined if all of the images or videos in the database have been learned. If not, the process repeats at block 400. Otherwise, content is retrieved from the database at block 450 by associating certain transform codes or groups of transform codes with certain content.

### User Query

In one embodiment, a user is able to query the database of images and videos by either submitting an input image for retrieval or specifying a type of content to be retrieved. If the 5 type of content to be retrieved is specified, then images with sets of compression patterns associated with the content are retrieved. An example of this embodiment of the present invention is shown in Figure 5.

At block 500, the unsupervised learning method examines the compression patterns of 10 all of the images or videos in the database. At block 510, all of the compression patterns are stored in a dictionary. At block 520, certain groups of compression patterns are associated with certain types of content in the dictionary. At block 530, a user submits a text query for content (i.e., a dog). At block 540 all of the groups of compression patterns associated with the content (i.e., the dog) are found in the dictionary. At block 550, all of the images or videos in the 15 database that have compression patterns that match the compression patterns for the content are retrieved.

In another embodiment, a user searches the database by providing an input image. When the input image is submitted, a set of compression patterns (e.g., transform codes) is 20 obtained for the input image and matched with the sets of transform codes for the images in the database. To determine if the transform codes for the input and database images match closely enough, an optimal Bayes decision rule is used by one embodiment of the present invention.

Figure 6 is a block diagram of an embodiment of the present invention that uses an image-based user query. An image query 600 is obtained from a user. The image query is designed to retrieve image 610. Image 610 based on the query is transferred to compressed domain 620 where it first is passed to a dictionary of transform codes 630. Next, multiple 5 descriptions 640 are used to support query refinement. Next the information is passed to a database 650 and the retrieved image 660 is obtained.

Figure 7 is a flowchart showing a process according to an embodiment of the present invention. At block 700, the unsupervised learning method examines the compression patterns 10 of all of the images or videos in the database. At block 710, all of the compression patterns are stored in a dictionary. At block 720, a user submits an image as a query for content. At block 730 the input image is compressed. At block 740, the input image is subdivided into blocks. At block 750, the blocks are examined for their compression patterns.

15 At block 760, the compression patterns for an image or video in the database is obtained. At block 770, Bayes decision rule is used to determine whether the patterns in the database image or video are matched closely enough to the patterns in the input image. If not, then at block 780, the next image or video is obtained and block 760 repeats. Otherwise, at block 790, the database image or video is selected and block 760 repeats.

20

#### Gaussian Latent Variable Modeling

In one embodiment, the unsupervised learning method uses Gaussian latent variable modeling. Gaussian latent variable modeling assumes that the high-dimensional observed space

( $\chi$ ) is generated from a low-dimensional process defined by linear transformation latent variables, or hidden causes ( $\dot{Z}$ ), plus an additive noise ( $\mu$ ):

$$\chi = T\dot{Z} + \mu + \varepsilon$$

where the columns of  $T$  are the basis functions, and the latent variables and noise are defined as:

5  $\dot{Z} \sim N(0, I)$   $\varepsilon \sim N(0, \Psi)$

The observation vectors are also normally distributed.

$$\chi \sim N(\mu, TT^T + \Psi)$$

The goal is to find the  $T$  and  $\Psi$  that best model the covariance structure of  $\chi$ .

10 EM Learning of Model Parameters

In one embodiment, the learning process according to the present invention comprises an E step and an M step. The E step computes the moments  $E[z|x_i, w_j]$  and  $E[zz'|x_i, w_j]$  for all data points  $i$  and mixture components  $j$  given the current parameter values  $\Lambda_j$ , and  $\Psi_j$ .

15

$$h_{ij} = \pi_j N(x_i - \mu_j, \Lambda_j, \Lambda_j' + \Psi_j)$$

$$E[w_j Z | x_i] = h_{ij} \beta_j (x_i - \mu_j)$$

$$\beta_j \equiv \Lambda_j' (\Psi_j + \Lambda_j \Lambda_j')^{-1}$$

$$E[w_j z z' | x_i] = h_{ij} (I - \beta_j \Lambda_j + \beta_j (x_i - \mu_j) (x_i - \mu_j)' \beta_j')$$

20

The M step results in the following update equations for the parameters:

$$[\Lambda_j^{new} \mu_j^{new}] = (\sum_i h_y x_i E[\tilde{z}|x_i, w_j] (\sum_i h_y E[\tilde{z}|x_i, w_j])^{-1}$$

$$\tilde{\Psi}_j^{new} = \frac{1}{n} diag \{ \sum_y h_y (x_i - \tilde{\Lambda}_j^{new} E[\tilde{z}|x_i, w_j]) x_i' \}$$

5                     $\mu_j^{new} = \frac{1}{n} \sum_{i=1}^n h_y$

### Universal Statistical Coding

The Expectation Maximization (EM) algorithm is used to learn the parameters of a mixture of Gaussian linear subspaces (our dictionary of transform codes) for each image band.

- 10         The final step assigns an image block to the mixture component yielding the lowest reconstruction error. Different bit allocation matrices are learned for each transform basis, and the quantized coefficients are entropy encoded using an arithmetic encoder. Bayesian inference methods may be applied for image matching.

15         Coefficient Matching via Bayesian Evidential Reasoning

Figure 8 illustrates coefficient matching using Bayesian evidential reasoning. At the top level 800 a hypothesis  $H$  having the greatest probability given all the accumulated evidence is made. Level 820 represents some images that might exist in a database such as desert scenes

- 20         821, people 822, and water scenes 823. Here the belief in Hypothesis  $H$ , taking Evidence  $\bar{E}$

into Account is tested. From level 820 to 840  $E_i$  is evidence concerning hypothesis  $H_i$  from information source  $S_i$ , used to generate luminance nodes 841 – 843 and chrominance nodes up to n.

## 5 Bayesian Reasoning:

Uses an “Odds – Likelihood Ratio” formulation of Bayes’ rule, and assumes that evidence i are statistically

10

Prior

$$O(H) = P(H) /$$

Likelihood

$$i|H) = P(E_i|H) / P(E_i|\sim H)$$

Posteri

$$O(H|1, \dots, E_n) = O(H) \prod_{i=1}^n L(i|H)$$

15

A Posteriori Probability

$$P(H|1, \dots, E_n) = \frac{O(H|E_1, \dots, E_n)}{1 + O(H|E_1, \dots, E_n)}$$

## Experimental Results

### Experiment 1 – WUTC vs. Universal Statistical Coder

20

(VisTex) collection images were subdivided into 9 128x128 images using 5 for training, and 4 for testing. Matching only code usage histograms, and compressing to 0.4 bpp.

WUTC	78%	Correct Retrieval
UStatC	95%	Correct Retrieval

## Experiment 2 – Uncompressed vs. Compressed Domain Matching

- 5 real world image classes: birds, deserts, flowers, people, and water scenes, with 10 images/class.

Color Histograms	76%	Correct, Full Decompression
Level Zero Histograms	72%	Correct, (<35%) Decompression
Matching 1 <sup>st</sup> 5 Coeffs.	94%	Correct, Partial Decompression

## Object Detection

- This research addresses the problem of object detection in complex scenes, where the focus will  
15 be on detecting sub-classes of animals from still imagery.

The challenges are:

- image content may not be unique.
  - objects may be partially occluded or have part that blend with the background.
  - high degree of within class variation (e.g., type or breed, color, size, etc.)
  - variations due to background clutter, illumination, translations, scale, and 3D pose.

### System Block Diagram

A block diagram of a system for object detection is illustrated in Figure 9. Database 900 stores a plurality of images. In each image salient regions are located at multiple scales. These 5 become the searchable objects. These regions are provided to the learning system 920 and passed through a Bayes decision rule module 940.

### Object Recognition by Probabilistic Appearance Modeling

10 Using a mixture of factor analyzers framework (MFA), the clustering and dimensionality reduction steps are performed simultaneously within a maximum-likelihood framework.

#### Approach:

- Convert RGB to YCrCB) luminance and chrominance bands.
- 15 ● For each training image, find and extract salient subregions (8x8) within each band and at multiple scales.
- Learn a collection of subspaces within an MFA framework.
- The MFA model explicitly estimates the probability density of the class over the pattern space.
- 20 ● Bayes decision rule is used to decide the class.

## Detection of Salient Points

Interest points are local features at which the signal changes two dimensionally. The following matrix is computed in a local neighborhood around the point:

5

$$C = [\sum I_x \sum I_x I_y] \quad \text{compute the eigenvalues of } C$$

$$[\sum I_y \sum I_x^2] \quad \text{and test if } \lambda_1 > \lambda_2 > \text{constant}$$

## Comparison of Methods

10

100 images (50 per class) were used for training, and the same number for testing.

Confusion Matrix		MFA Local Appearance Modeling	
		Cats	Dogs
Cats	48	2	
Dogs	6	44	

92%

Correct

Confusion Matrix		Perceptual Color Histograms	
		Cats	Dogs
Cats	41	9	
Dogs	7	43	

84%

Correct

(equation)

Embodiment of Computer Execution Environment (Hardware)

An embodiment of the invention can be implemented as computer software in the form  
5 of computer readable program code executed in a general purpose computing environment such  
as environment 1000 illustrated in Figure 10, or in the form of bytecode class files executable  
within a Java™ run time environment running in such an environment, or in the form of  
bytecodes running on a processor (or devices enabled to process bytecodes) existing in a  
distributed environment (e.g., one or more processors on a network). A keyboard 1010 and  
10 mouse 1011 are coupled to a system bus 1018. The keyboard and mouse are for introducing user  
input to the computer system and communicating that user input to central processing unit  
(CPU) 1013. Other suitable input devices may be used in addition to, or in place of, the mouse  
1011 and keyboard 1010. I/O (input/output) unit 1019 coupled to bi-directional system bus  
1018 represents such I/O elements as a printer, A/V (audio/video) I/O, etc.

15

Computer 1001 may include a communication interface 1020 coupled to bus 1018.  
Communication interface 1020 provides a two-way data communication coupling via a network  
link 1021 to a local network 1022. For example, if communication interface 1020 is an  
integrated services digital network (ISDN) card or a modem, communication interface 1020  
20 provides a data communication connection to the corresponding type of telephone line, which  
comprises part of network link 1021. If communication interface 1020 is a local area network  
(LAN) card, communication interface 1020 provides a data communication connection via  
network link 1021 to a compatible LAN. Wireless links are also possible. In any such

implementation, communication interface 1020 sends and receives electrical, electromagnetic or optical signals which carry digital data streams representing various types of information.

Network link 1021 typically provides data communication through one or more  
5 networks to other data devices. For example, network link 1021 may provide a connection  
through local network 1022 to local server computer 1023 or to data equipment operated by ISP  
1024. ISP 1024 in turn provides data communication services through the world wide packet  
data communication network now commonly referred to as the "Internet" 1025. Local network  
1022 and Internet 1025 both use electrical, electromagnetic or optical signals which carry digital  
10 data streams. The signals through the various networks and the signals on network link 1021  
and through communication interface 1020, which carry the digital data to and from computer  
1000, are exemplary forms of carrier waves transporting the information.

Processor 1013 may reside wholly on client computer 1001 or wholly on server 1026 or  
15 processor 1013 may have its computational power distributed between computer 1001 and  
server 1026. Server 1026 symbolically is represented in Figure 10 as one unit, but server 1026  
can also be distributed between multiple "tiers". In one embodiment, server 1026 comprises a  
middle and back tier where application logic executes in the middle tier and persistent data is  
obtained in the back tier. In the case where processor 1013 resides wholly on server 1026, the  
20 results of the computations performed by processor 1013 are transmitted to computer 1001 via  
Internet 1025, Internet Service Provider (ISP) 1024, local network 1022 and communication  
interface 1020. In this way, computer 1001 is able to display the results of the computation to a  
user in the form of output.

Computer 1001 includes a video memory 1014, main memory 1015 and mass storage 1012, all coupled to bi-directional system bus 1018 along with keyboard 1010, mouse 1011 and processor 1013.

As with processor 1013, in various computing environments, main memory 1015 and mass storage 1012, can reside wholly on server 1026 or computer 1001, or they may be distributed between the two. Examples of systems where processor 1013, main memory 1015, and mass storage 1012 are distributed between computer 1001 and server 1026 include the thin-client computing architecture developed by Sun Microsystems, Inc., the palm pilot computing device and other personal digital assistants, Internet ready cellular phones and other Internet computing devices, and in platform independent computing environments, such as those which utilize the Java technologies also developed by Sun Microsystems, Inc.

The mass storage 1012 may include both fixed and removable media, such as magnetic, optical or magnetic optical storage systems or any other available mass storage technology. Bus 1018 may contain, for example, thirty-two address lines for addressing video memory 1014 or main memory 1015. The system bus 1018 also includes, for example, a 32-bit data bus for transferring data between and among the components, such as processor 1013, main memory 1015, video memory 1014 and mass storage 1012. Alternatively, multiplex data/address lines may be used instead of separate data and address lines.

20

In one embodiment of the invention, the processor 1013 is a microprocessor manufactured by Motorola, such as the 680X0 processor or a microprocessor manufactured by Intel, such as the 80X86, or Pentium processor, or a SPARC microprocessor from Sun Microsystems, Inc. However, any other suitable microprocessor or microcomputer may be

utilized. Main memory 1015 is comprised of dynamic random access memory (DRAM). Video memory 1014 is a dual-ported video random access memory. One port of the video memory 1014 is coupled to video amplifier 1016. The video amplifier 1016 is used to drive the cathode ray tube (CRT) raster monitor 1017. Video amplifier 1016 is well known in the art and may be  
5 implemented by any suitable apparatus. This circuitry converts pixel data stored in video memory 1014 to a raster signal suitable for use by monitor 1017. Monitor 1017 is a type of monitor suitable for displaying graphic images.

Computer 1001 can send messages and receive data, including program code, through  
10 the network(s), network link 1021, and communication interface 1020. In the Internet example, remote server computer 1026 might transmit a requested code for an application program through Internet 1025, ISP 1024, local network 1022 and communication interface 1020. The received code may be executed by processor 1013 as it is received, and/or stored in mass storage 1012, or other non-volatile storage for later execution. In this manner, computer 1000 may  
15 obtain application code in the form of a carrier wave. Alternatively, remote server computer 1026 may execute applications using processor 1013, and utilize mass storage 1012, and/or video memory 1015. The results of the execution at server 1026 are then transmitted through Internet 1025, ISP 1024, local network 1022 and communication interface 1020. In this example, computer 1001 performs only input and output functions.

20

Application code may be embodied in any form of computer program product. A computer program product comprises a medium configured to store or transport computer readable code, or in which computer readable code may be embedded. Some examples of

computer program products are CD-ROM disks, ROM cards, floppy disks, magnetic tapes, computer hard drives, servers on a network, and carrier waves.

The computer systems described above are for purposes of example only. An  
5 embodiment of the invention may be implemented in any type of computer system or  
programming or processing environment.

Thus, a method and apparatus for indexing images and video is described in conjunction  
with one or more specific embodiments. The invention is defined by the claims and their full  
10 scope of equivalents.

7,